

## METHYLATION ALTERED DNA SEQUENCES AS MARKERS ASSOCIATED WITH HUMAN CANCER

### 5 Technical Field of the Invention

The present invention relates to novel human DNA sequences that exhibit altered methylation patterns (hypermethylation or hypomethylation) in cancer patients. These novel methylation-altered DNA sequences are useful as diagnostic, prognostic and therapeutic markers for human cancer.

10

### Background of the Invention

The identification of early genetic changes in tumorigenesis is a primary focus in molecular cancer research. Characterization of the nature and pattern of cancer-associated genetic alterations will allow for early detection, diagnosis and treatment of cancer. Such 15 genetic alterations in vertebrates fall generally into one of three categories: gain or loss of genetic material; mutation of genetic material; or methylation at cytosine residues in CpG dinucleotides within "CpG islands." Among these, DNA methylation is uniquely reversible, and changes in methylation state are known to affect gene expression (e.g., transcriptional initiation of genes where CpG islands located at or near the promoter region) or genomic 20 stability.

25 *Methylation of CpG dinucleotides within CpG islands.* DNA, in higher order eukaryotic organisms, is methylated only at cytosine residues located 5' to guanosine residues in CpG dinucleotides. This covalent modification of the C-5 position of the cytosine base by the enzyme DNA (cytosine-5)-methyltransferase results in the formation of 5-methylcytosine (5-mCyt), and gives this base unique properties (e.g., susceptibility to undergo spontaneous deamination). This enzymatic conversion is the only epigenetic modification of DNA known to exist in vertebrates, and is essential for normal embryonic development (Bird, A.P., *Cell* 70:5-8, 1992; Laird & Jaenisch, *Human Molecular Genetics* 3:1487-1495, 1994; Li et al., *Cell* 69:915-926, 1992).

30 The presence of 5-mCyt at CpG dinucleotides has resulted in the 5-fold depletion of this sequence in the genome during the course of vertebrate evolution (Schroderet & Gartler, *Proc. Nat. Acad. Sci. USA* 89:957-961, 1992), presumably due to spontaneous deamination of 5-mCyt to Thymidine. Certain areas of the genome, however, do not show such depletion, and are referred to as "CpG islands" (Bird, A.P., *Nature* 321:209-213, 1986; Gardiner- 35 Garden & Frommer, *J. Mol. Biol.* 196:261-282, 1987). These CpG islands comprise only approximately 1% of the vertebrate genome, yet account for about 15% of the total number of genomic CpG dinucleotides (Antequera & Bird, *Proc. Nat. Acad. Sci. USA* 90:11995-11999, 1993). CpG islands contain the expected (i.e., the non-evolutionarily depleted) frequency of

CpGs (with an Observed/Expected Ratio<sup>1</sup> >0.6), are GC-rich (with a GC Content<sup>2</sup> >0.5) and are typically between about 0.2 to about 1 kb in length.

*Methylation within CpG islands affects gene expression.* CpG islands are located upstream of many housekeeping and tissue-specific genes, but may also extend into gene coding regions (Cross & Bird, *Current Opinions in Genetics and Development* 5:309-314, 1995; Larsen et al., *Genomics* 13:1095-1107, 1992). The methylation of cytosines within CpG islands in somatic tissues is believed to affect gene expression. Methylation has been inversely correlated with gene activity and may lead to decreased gene expression by a variety of mechanisms including inhibition of transcription initiation (Bird, A.P., *Nature* 321:209-213, 1986; Delgado et al., *EMBO Journal* 17:2426-2435, 1998), disruption of local chromatin structure (Counts & Goodman, *Molecular Carcinogenesis* 11:185-188, 1994; Antequera et al., *Cell* 62:503-514, 1990), and recruitment of proteins that interact specifically with methylated sequences and thereby directly or indirectly prevent transcription factor binding (Bird, A.P., *Cell* 70:5-8, 1992; Counts & Goodman, *Molecular Carcinogenesis* 11:185-188, 1994; Cedar, H., *Cell* 53:3-4, 1988). Many studies have demonstrated the effect of methylation of CpG islands on gene expression (e.g., the CDKN2A/p16 gene; Gonzalez-Zulueta et al., *Cancer Research* 55:4531-4535, 1995), but most CpG islands on autosomal genes remain unmethylated in the germline, and methylation of these islands is usually independent of gene expression. Tissue-specific genes are typically unmethylated in the respective target organs but are methylated in the germline and in non-expressing adult tissues, while CpG islands of constitutively expressed housekeeping genes are normally unmethylated in the germline and in somatic tissues.

*Methylation within CpG islands affects the expression of genes involved in cancer.* Data from a group of studies show the presence of altered methylation in cancer cells relative to non-cancerous cells. These studies show not only alteration of the overall genomic levels of DNA methylation, but also changes in the distribution of methyl groups. For example, abnormal methylation of CpG islands that are associated with tumor suppressor genes or oncogenes within a cell may cause altered gene expression. Such altered gene expression may provide a population of cells with a selective growth advantage and thereby result in selection of these cells to the detriment of the organism (*i.e.*, cancer).

*Insufficient correlative data.* Unfortunately, the mere knowledge of the basic existence of altered methylation of CpG dinucleotides within CpG islands of cancer cells relative to normal cells, or of the fact that in particular instances such methylation changes result in altered gene expression (or chromatin structure or stability), is inadequate to allow for effective diagnostic, prognostic and therapeutic application of this knowledge. This is

<sup>1</sup> Calculated as: [number of CpG sites / (number of C bases X number of G bases)] X band length for each fragment.

<sup>2</sup> Calculated as: (number of C bases + number of G bases) / band length for each fragment.

because only a limited number of CpG islands have been characterized, and thus there is insufficient knowledge, as to which particular CpG islands, among many, are actually involved in, or show significant correlation with cancer or the etiology thereof. Moreover, complex methylation patterns, involving a plurality of methylation-altered DNA sequences, 5 including those that may have the sequence composition to qualify as CpG islands, may exist in particular cancers.

Therefore there is a need in the art to identify and characterize specific methylation altered DNA sequences, and to correlate them with cancer to allow for their diagnostic, prognostic and therapeutic application.

10

### **Summary of the Invention**

The present invention provides for a diagnostic or prognostic assay for cancer, comprising: obtaining a tissue sample from a test tissue; performing a methylation assay on DNA derived from the tissue sample, wherein the methylation assay determines the 15 methylation state of a CpG dinucleotide within a DNA sequence of the DNA, and wherein the DNA sequence is a sequence selected from the group consisting of sequences of SEQ ID NOS:1-103, sequences having a nucleotide sequence at least 90% identical to sequences of SEQ ID NOS:1-103, CpG island sequences associated with sequences of SEQ ID NOS:1-103, CpG island sequences associated with sequences having a nucleotide sequence at least 90% identical to sequences of SEQ ID NOS:1-103, and combinations thereof, wherein the 20 CpG island sequence associated with the sequence of the particular SEQ ID NO is that contiguous sequence of genomic DNA that encompasses at least one nucleotide of the particular SEQ ID NO sequence, and satisfies the criteria of having both a frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6, and a GC Content >0.5; 25 and determining a diagnosis or prognosis based, at least in part, upon the methylation state of the CpG dinucleotide within the DNA sequence. Preferably, the DNA sequence is a sequence selected from the group consisting of CpG island sequences associated with sequences of SEQ ID NOS:1-103, CpG island sequences associated with sequences having a nucleotide sequence at least 90% identical to sequences of SEQ ID NOS:1-103, and 30 combinations thereof. Preferably, the DNA sequence is a sequence selected from the group consisting of CpG island sequences associated with sequences of SEQ ID NOS: 2, 4, 6, 7, 9-16, 19, 20, 22-33, 35-43, 48, 51-55, 59, 60, 64, 71, 76, 78-81, 84 and 87-90, and combinations thereof. Preferably, the methylation assay procedure is selected from the group 35 consisting of MethylLight, MS-SnuPE (methylation-sensitive single nucleotide primer extension), MSP (methylation-specific PCR), MCA (methylated CpG island amplification), COBRA (combined bisulfite restriction analysis), and combinations thereof. Preferably, the methylation state of the CpG dinucleotide within the DNA sequence is that of hypermethylation, hypomethylation or normal methylation. Preferably, the cancer is selected

from the group consisting of bladder cancer, prostate cancer, colon cancer, lung cancer, renal cancer, leukemia, breast cancer, uterine cancer, astrocytoma, glioblastoma, and neuroblastoma. Preferably, the cancer is bladder cancer, or prostate cancer.

- The present invention further provides a kit useful for the detection of a methylated
- 5 CpG-containing nucleic acid comprising a carrier means containing one or more containers comprising: a container containing a probe or primer which hybridizes to any region of a sequence selected from the group consisting of SEQ ID NOS:1-103, and sequences having a nucleotide sequence at least 90% identical to sequences of SEQ ID NOS:1-103; and additional standard methylation assay reagents required to affect detection of methylated
- 10 CpG-containing nucleic acid based on the probe or primer. Preferably, the additional standard methylation assay reagents are standard reagents for performing a methylation assay from the group consisting of MethylLight, MS-SNuPE, MSP, MCA, COBRA, and combinations thereof. Preferably, the probe or primer comprises at least about 12 to 15 nucleotides of a sequence selected from the group consisting of SEQ ID NOS:1-103, and
- 15 sequences having a nucleotide sequence at least 90% identical to sequences of SEQ ID NOS:1-103.

The present invention further provides an isolated nucleic acid molecule comprising a methylated or unmethylated polynucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:74, SEQ ID NO:76, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:86, SEQ ID NO:90, SEQ ID NO:92, SEQ ID NO:97, and SEQ ID NO:100. Preferably the nucleic acid is methylated. Preferably, the nucleic acid is unmethylated.

## Detailed Description of the Invention

### 30 Definitions:

“**GC Content**” refers, within a particular DNA sequence, to the [(number of C bases + number of G bases) / band length for each fragment].

“**Observed/Expected Ratio**” (“**O/E Ratio**”) refers to the frequency of CpG dinucleotides within a particular DNA sequence, and corresponds to the [number of CpG sites / (number of C bases X number of G bases)] X band length for each fragment.

“**CpG Island**” refers to a contiguous region of genomic DNA that satisfies the criteria of (1) having a frequency of CpG dinucleotides corresponding to an “Observed/Expected Ratio” >0.6, and (2) having a “GC Content” >0.5. CpG islands are

typically, but not always, between about 0.2 to about 1 kb in length. A CpG island sequence associated with a particular SEQ ID NO sequence of the present invention is that contiguous sequence of genomic DNA that encompasses at least one nucleotide of the particular SEQ ID NO sequence, and satisfies the criteria of having both a frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6), and a GC Content >0.5.

5        “**Methylation state**” refers to the presence or absence of 5-methylcytosine (“5-mCyt”) at one or a plurality of CpG dinucleotides within a DNA sequence.

10      “**Hypermethylation**” refers to the methylation state corresponding to an *increased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

15      “**Hypomethylation**” refers to the methylation state corresponding to a *decreased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

“**Methylation assay**” refers to any assay for determining the methylation state of a CpG dinucleotide within a sequence of DNA.

20      “**MS-AP-PCR**” (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognized technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al., *Cancer Research* 57:594-599, 1997.

25      “**MethyLight**” refers to the art-recognized fluorescence-based real-time PCR technique described by Eads et al., *Cancer Res.* 59:2302-2306, 1999.

“**Ms-SNuPE**” (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognized assay described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

30      “**MSP**” (Methylation-specific PCR) refers to the art-recognized methylation assay described by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by US Patent No. 5,786,146.

“**COBRA**” (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997.

“**MCA**” (Methylated CpG Island Amplification) refers to the methylation assay described by Toyota et al., *Cancer Res.* 59:2307-12, 1999, and in WO 00/26401A1.

### 35      **Overview**

The present invention provides for 103 DNA sequences (*i.e.*, “marker sequences”) having distinct methylation patterns in cancer, as compared to normal tissue. These methylation-altered DNA sequence embodiments correspond to 103 DNA fragments isolated

from bladder and prostate cancer patients, and in many instances, represent novel sequences not found in the GenBank database. *None* of the instant sequence embodiments have previously been characterized with respect to their methylation pattern in human cancers including, but not limited to, those of the bladder and prostate. The significance of such 5 methylation patterns lies in the value of altered fragments as potential prognostic, diagnostic and therapeutic markers in the treatment of human cancers.

### **Identification of Methylation-altered Marker Sequences in Genomic DNA**

The MS.AP-PCR technique was used to scan the genomes of bladder or prostate 10 cancer patients for DNA methylation changes relative to normal individuals, because the pattern is known to be highly conserved. A total of 103 DNA sequence embodiments (methylation-altered DNA sequences; "marker sequences") were isolated and characterized as having distinct methylation patterns in cancer, as compared to normal tissue.

**Methods for the Identification of Marker Sequences in Genomic DNA.** There are a 15 variety of art-recognized genome scanning methods that have been used to identify altered methylation sites in cancer cells. For example, one method involves restriction landmark genomic scanning (Kawai et al., *Mol. Cell. Biol.* 14:7421-7427, 1994), another involves MCA (methylated CpG island amplification; Toyota et al., *Cancer Res.* 59:2307-12, 1999), and yet another involves MS.AP-PCR (Methylation-Sensitive Arbitrarily-Primed Polymerase 20 Chain Reaction; Gonzalgo et al., *Cancer Res.* 57:594-599, 1997), which allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides. The MS.AP-PCR technique used in the present invention is a rapid and efficient method to screen ("scan") for altered methylation patterns in genomic DNA and to isolate specific sequences associated with these changes.

Briefly, genomic DNA from the tissue of bladder or prostate cancer patients was 25 prepared using standard, art-recognized methods. Restriction enzymes (*e.g.*, HpaII) with different sensitivities to cytosine methylation in their recognition sites were used to digest these genomic DNAs prior to arbitrarily primed PCR amplification with GC-rich primers. Fragments that showed differential methylation (*e.g.*, *hypermethylation* or *hypomethylation*, 30 based on the methylation sensitivity of the restriction enzyme, or upon DNA sequence analysis or Ms-SNuPE analysis; Gonzalgo & Jones, *Nucleic Acids Res* 25:2529-2531, 1997) were cloned and sequenced after resolving the PCR products on high-resolution polyacrylamide gels. The cloned fragments were used as probes for Southern blot analysis to 35 confirm differential methylation of these regions in the tissue. Methods for DNA cloning, sequencing, PCR, high-resolution polyacrylamide gel resolution and Southern blot analysis are well known by those of ordinary skill in the relevant art.

**Results.** A total of 500 DNA fragments that underwent either hypermethylation (an increase in the level of methylation relative to normal) or hypomethylation (a decrease in the

level of methylation relative to normal) were isolated from the selected patients genomic DNA. A total of 178 of these fragments were sequenced, of which 103 were *novel* in that they corresponded to DNA loci whose methylation pattern had not previously been characterized. The corresponding sequences are disclosed as [SEQ ID NOS:1-103], wherein  
5 for certain sequences, the letter "n" refers to an undetermined nucleotide base.

**Novel marker sequences identified by MS.AP-PCR.** Table I shows an overall summary of methylation patterns and sequence data corresponding to the 103 DNA fragments identified by MS.AP-PCR. A total of 103 fragments were sequenced following identification as becoming either hypermethylated (gain of methylation; noted as having a  
10 hypermethylation pattern) or hypomethylated (loss of methylation; noted as having a hypomethylation pattern) relative to normal tissue. For the fragments of each category, the "Average GC Content" is shown, calculated as (number of C bases + number of G bases)/band length for each fragment, as well as the average Observed/Expected Ratio ("O/E Ratio"), calculated as [number of CpG sites/(number of C bases X number of G bases)] X  
15 band length for each fragment. Additionally, the percent of fragments that qualify as CpG islands is listed, and corresponds to the percentage of all fragments within each category that have sequence compositions that satisfy the criteria of having a "GC Content" >0.5 and an "O/E Ratio" >0.6.

Thus, of these 103 fragments identified by MS.AP-PCR, 60 showed hypermethylation (Table I, upper row; Table II, [SEQ ID NOS:1-60]) while 43 showed hypomethylation (Table I, lower row; Table II, [SEQ ID NOS:61-103]). Moreover, 55 (43 hypermethylated, and 12 hypomethylated) of the 103 fragments correspond to CpG islands (*i.e.*, fulfill the criteria of a GC content >0.5 and an Observed/Expected Ratio >0.6), whereas the other 48 (17 hypermethylated and 31 hypomethylated) fragments do not meet the criteria for CpG islands  
25 (see Table II).

**TABLE I. Summary of 103 DNA Fragments Identified by MS.AP-PCR**

DNA Fragment Type	Methylation Pattern (relative to normal)	Number of Fragments (103 total)	Average GC Content	Average O/E Ratio	Percent that correspond to CpG Islands
Hypermethylated Fragments	Hyper-methylation	60	0.54	0.72	72%
Hypomethylated Fragments	Hypo-methylation	43	0.52	0.48	28%

Table II shows a summary of methylation pattern and sequence data for each individual sequence embodiment ([SEQ ID NOS:1-103]), corresponding to the 103 DNA fragments identified by MS.AP-PCR. Data for the 103 fragments was divided into either hypermethylated ([SEQ ID NOS:1-60]) or hypomethylated ([SEQ ID NOS:61-103]) categories. Table II also lists, for each sequence embodiment, the corresponding "Fragment  
30

Name," fragment "Size" (in base pairs; "bp"), "GC Content," Observed/Expected Ratio ("O/E Ratio"), "Description" (*i.e.*, as a CpG island if criteria are met), "Inventor Initials" (IDCM = Isabel D.C. Markl, JC = Jonathan Cheng, GL = Gangning Liang, HF = Hualin Fu, YT = Yoshitaka Tomigahara), "Cancer Source," and "Chromosome Match" to the GenBank database. A dash ("–") indicates that no GenBank chromosome match existed, or that only a low-scoring partial match was found. Averages of the "GC Content" and "O/E Ratio," along with the percent of fragments that are CpG islands, are listed after the last member of both the hypermethylated and hypomethylated categories.

Therefore, the present invention provides for 103 DNA fragments and corresponding marker sequence embodiments (*i.e.*, methylation-altered DNA sequences) that are useful in cancer prognostic, diagnostic and therapeutic applications.

Additionally, at least 55 of these 103 sequences correspond to CpG islands (based on GC Content and O/E ration); namely [SEQ ID NOS:2, 4, 6, 7, 9-16, 19, 20, 22-33, 35-43, 48, 51-55, 59, 60, 64, 71, 76, 78-81, 84 and 87-90]. Thus, based on the fact that the methylation state of a portion of a given CpG island is generally representative of the island as a whole, the present invention further encompassed the novel use of the 55 CpG islands associated with [SEQ ID NOS:2, 4, 6, 7, 9-16, 19, 20, 22-33, 35-43, 48, 51-55, 59, 60, 64, 71, 76, 78-81, 84 and 87-90] in cancer prognostic, diagnostic and therapeutic applications, where a CpG island sequence associated with the sequence of a particular SEQ ID NO is that contiguous sequence of genomic DNA that encompasses at least one nucleotide of the particular SEQ ID NO sequence, and satisfies the criteria of having both a frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6, and a GC Content >0.5.

**TABLE II. Summary of MS.AP-PCR Fragments Sequenced**

Methylation Pattern	Fragment Name	Size (bp)	GC Content	O/E Ratio	Description	Inventor Initials	Cancer Source	Chromosome Matches	[SEQ ID NO]
Hyper-methylation category	11-1A	510	0.44	0.74		IDCM	Bladder	-	1
	14-3B	313	0.58	0.74	CpG Island	IDCM	Bladder	2	2
	18-2B	165	0.57	0.45		IDCM	Bladder	7	3
	24-1B	601	0.51	0.72	CpG Island	IDCM	Bladder		Xp11
	26-1B	801	0.48	0.56		IDCM	Bladder	-	5
	26-2C	204	0.50	0.63	CpG Island	IDCM	Bladder	-	6
	30-3D	205	0.55	1.25	CpG Island	IDCM	Bladder	14	7
	32-3E	597	0.57	0.10		IDCM	Bladder		8
	34-2B	500	0.62	0.66	CpG Island	IDCM	Bladder	20q12-13.1	9
	34-4B	343	0.70	0.81	CpG Island	IDCM	Bladder	-	10
	34-5D	291	0.62	0.96	CpG Island	IDCM	Bladder	9	11
	34-6A	266	0.64	0.93	CpG Island	IDCM	Bladder	-	12
	35-1C	553	0.64	0.63	CpG Island	IDCM	Bladder	-	13
	36-2D	156	0.60	0.58	CpG Island	IDCM	Bladder	10	14
	38-1A	300	0.70	0.80	CpG Island	IDCM	Bladder	10	15
	38-2B	196	0.56	0.89	CpG Island	IDCM	Bladder	15	16
	7-8E	299	0.59	0.39		IDCM	Bladder	17q21-22	17
	83-4B	363	0.54	0.49		IDCM	Bladder	-	18

Methylation Pattern	Fragment Name	Size (bp)	GC Content	O/E Ratio	Description	Inventor Initials	Cancer Source	Chromosome Matches	[SEQ ID NO]
	84-1D	322	0.55	0.90	CpG Island	IDCM	Bladder	7	19
	101-3E	255	0.57	0.83	CpG Island	IDCM	Bladder	17	20
	M1-5A	406	0.45	0.96		IDCM	Bladder	1	21
	U2-8E	210	0.56	0.61	CpG Island	IDCM	Bladder	2	22
	U12-1A	310	0.56	0.81	CpG Island	IDCM	Bladder	2	23
	U7-4A	305	0.59	0.80	CpG Island	IDCM	Bladder	-	24
	NU9-5A	379	0.67	0.83	CpG Island	JC	Bladder	-	25
	3-17-8-B	625	0.48	0.72	CpG Island	GL	Bladder	18	26
	4-10-4-A	499	0.55	0.30	CpG Island	GL	Bladder	7	27
	1-1-1-A	561	0.58	0.98	CpG Island	GL	Bladder	20	28
	3-17-8-A	717	0.50	0.68	CpG Island	GL	Bladder	17	29
	G145-H	280	0.50	1.10	CpG Island	GL	Bladder	11	30
	1-1-1-D	270	0.50	0.60	CpG Island	GL	Bladder	2	31
	1-1-1-C	347	0.65	1.25	CpG Island	GL	Bladder	-	32
	G178-A	342	0.55	0.85	CpG Island	GL	Bladder	-	33
	34-A	370	0.62	0.44		HF	Prostate	-	34
	34-D	213	0.53	0.74	CpG Island	HF	Prostate	2	35
	35-D	173	0.56	0.66	CpG Island	HF	Prostate	3	36
	36-A	369	0.67	0.70	CpG Island	HF	Prostate	-	37
	40-A	123	0.60	1.16	CpG Island	HF	Prostate	-	38
	91-1	450	0.64	0.86	CpG Island	YT	Bladder	5 or 16q24.3	39
	93-2	593	0.51	0.68	CpG Island	YT	Bladder	Xp11	40
	93-3	457	0.52	0.94	CpG Island	YT	Bladder	Xp22.1-22.3	41
	94-8	211	0.66	0.96	CpG Island	YT	Bladder	-	42
	95-5	141	0.63	0.79	CpG Island	YT	Bladder	14	43
	97-5	559	0.56	0.40		YT	Bladder	-	44
	98-1	433	0.46	0.96		YT	Bladder	1	45
	100-1	487	0.59	0.58		YT	Bladder	14	46
	100-2	403	0.60	0.47		YT	Bladder	3	47
	100-6	155	0.57	0.99	CpG Island	YT	Bladder	20	48
	4-2	256	0.57	0.40		YT	Bladder	7	49
	5-8	224	0.47	0.96		YT	Bladder	5	50
	6-4	313	0.70	0.82	CpG Island	YT	Bladder	-	51
	7-6	385	0.70	0.88	CpG Island	YT	Bladder	-	52
	13-3	307	0.59	0.89	CpG Island	YT	Bladder	10	53
	15-2	182	0.62	0.92	CpG Island	YT	Bladder	13	54
	23-2	523	0.54	0.87	CpG Island	YT	Bladder	Xp22.1-22.3	55
	39-2	795	0.46	0.64		YT	Bladder	13	56
	40-2	438	0.62	0.51		YT	Bladder	10	57
	41-3	611	0.47	0.70		YT	Bladder	18	58
	105-4	291	0.58	0.71	CpG Island	YT	Bladder	5	59
	107-8	226	0.53	0.96	CpG Island	YT	Bladder	11	60
AVERAGE			0.54	0.72	72% islands				
Hypo-methylation category	14-2B	580	0.55	0.51		IDCM	Bladder	2	61
	16-1B	633	0.56	0.39		IDCM	Bladder	-	62
	18-1B	703	0.45	0.35		IDCM	Bladder	17	63
	19-1B	420	0.66	0.87	CpG Island	IDCM	Bladder	-	64
	20-1B	496	0.61	0.59		IDCM	Bladder	-	65
	21-2C	637	0.60	0.33		IDCM	Bladder	9q34	66
	29-1A	595	0.55	0.27		IDCM	Bladder	Xp11.23	67
	29-2B	580	0.47	0.77		IDCM	Bladder	-	68
	32-1A	589	0.59	0.48		IDCM	Bladder	-	69
	34-1B	450	0.42	0.46		IDCM	Bladder	-	70
	34-3B	432	0.70	0.61	CpG Island	IDCM	Bladder	-	71

Methylation Pattern	Fragment Name	Size (bp)	GC Content	O/E Ratio	Description	Inventor Initials	Cancer Source	Chromosome Matches	[SEQ ID NO]
	32-2B	748	0.47	0.24	CpG Island	IDCM	Bladder	2	72
	32-4B	599	0.57	0.15		IDCM	Bladder	20q12-13.1	73
	32-5B	614	0.58	0.20		IDCM	Bladder	-	74
	33-1A	552	0.54	0.32		IDCM	Bladder	10	75
	5-1E	501	0.61	1.04		IDCM	Bladder	-	76
	6-1A	826	0.55	0.36		IDCM	Bladder	22q13.32-13.33	77
	7-5D	433	0.59	0.85		IDCM	Bladder	5	78
	8-7C	424	0.58	0.83		IDCM	Bladder	5	79
	30-6D	285	0.63	0.72		IDCM	Bladder	1	80
	66-2E	401	0.54	0.82		IDCM	Bladder	16	81
	78-1C	268	0.54	0.41		IDCM	Bladder	-	82
	97-2E	989	0.53	0.16		IDCM	Bladder	-	83
	M1-8C	250	0.64	0.99		IDCM	Bladder	-	84
	M2-5A	402	0.50	0.45		IDCM	Bladder	5	85
	M1-4P	595	0.43	0.41		IDCM	Bladder	-	86
	M12-10A	304	0.53	0.76		IDCM	Bladder	7	87
	M12-12C	296	0.51	0.64		IDCM	Bladder	17	88
	M2-8M	220	0.67	0.62		IDCM	Bladder	6q27	89
	NU4-3A	273	0.63	1.02		JC	Bladder	-	90
	NU5-2A	361	0.44	0.73		JC	Bladder	6q14.3-15	91
	88-5	462	0.62	0.39		YT	Bladder	-	92
	90-1	591	0.66	0.45		YT	Bladder	19	93
	91-3	279	0.58	0.45		YT	Bladder	5 or 16q24.3	94
	91-4	351	0.55	0.30		YT	Bladder	18q23	95
	91-7	171	0.61	0.59		YT	Bladder	11	96
	89-3	743	0.55	0.43		YT	Bladder	-	97
	94-2	589	0.53	0.41		YT	Bladder	22q13.31-13.32	98
	94-3	538	0.53	0.49		YT	Bladder	5 or 18	99
	94-4	486	0.61	0.57		YT	Bladder	-	100
	94-5	450	0.60	0.45		YT	Bladder	1p36.2-36.3	101
	94-6	292	0.58	0.32		YT	Bladder	8 or 9	102
	96-4	395	0.63	0.54		YT	Bladder	9	103
AVERAGE		0.52	0.48	28% islands					

**Diagnostic and Prognostic Assays for Cancer.** The present invention provides for diagnostic and prognostic cancer assays based on determination of the methylation state of one or more of the disclosed 103 methylation-altered DNA sequence embodiments.

Typically, such assays involve obtaining a tissue sample from a test tissue, performing a methylation assay on DNA derived from the tissue sample, and making a diagnosis or prognosis based thereon.

The methylation assay is used to determine the methylation state of one or a plurality of CpG dinucleotide within a DNA sequence of the DNA sample. According to the present invention, possible methylation states include *hypermethylation* and *hypomethylation*, relative to a normal state (*i.e.*, non-cancerous control state). Hypermethylation and hypomethylation refer to the methylation states corresponding to an *increased* or *decreased*, respectively,

presence 5-methylcytosine ("5-mCyt") at one or a plurality of CpG dinucleotides within a DNA sequence of the test sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

A diagnosis or prognosis is based, at least in part, upon the determined methylation state of the sample DNA sequence compared to control data obtained from normal, non-cancerous tissue.

**Methylation Assay Procedures.** Various methylation assay procedures are known in the art, and can be used in conjunction with the present invention. These assays allow for determination of the methylation state of one or a plurality of CpG dinucleotides (e.g., CpG islands) within a DNA sequence. Such assays involve, among other techniques, DNA sequencing of bisulfite-treated DNA, PCR (for sequence-specific amplification), Southern blot analysis, use of methylation-sensitive restriction enzymes, etc.

For example, genomic sequencing has been simplified for analysis of DNA methylation patterns and 5-methylcytosine distribution by using bisulfite treatment (Frommer et al., *Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA is used, e.g., the method described by Sadri & Hornsby (*Nucl. Acids Res.* 24:5058-5059, 1996), or COBRA (Combined Bisulfite Restriction Analysis) (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997).

**COBRA.** COBRA analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific gene loci in small amounts of genomic DNA (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997). Briefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated DNA. Methylation-dependent sequence differences are first introduced into the genomic DNA by standard bisulfite treatment according to the procedure described by Frommer et al. (*Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). PCR amplification of the bisulfite converted DNA is then performed using primers specific for the interested CpG islands, followed by restriction endonuclease digestion, gel electrophoresis, and detection using specific, labeled hybridization probes. Methylation levels in the original DNA sample are represented by the relative amounts of digested and undigested PCR product in a linearly quantitative fashion across a wide spectrum of DNA methylation levels. In addition, this technique can be reliably applied to DNA obtained from microdissected paraffin-embedded tissue samples. Typical reagents (e.g., as might be found in a typical COBRA-based kit) for COBRA analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridization oligo; control hybridization oligo; kinase labeling kit for oligo probe; and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery regents or kit (e.g.,

precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Preferably, assays such as "MethyLight" (a fluorescence-based real-time PCR technique) (Eads et al., *Cancer Res.* 59:2302-2306, 1999), Ms-SNuPE (Methylation-sensitive Single Nucleotide Primer Extension) reactions (Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997), methylation-specific PCR ("MSP"; Herman et al., *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; US Patent No. 5,786,146), and methylated CpG island amplification ("MCA"; Toyota et al., *Cancer Res.* 59:2307-12, 1999) are used alone or in combination with other of these methods.

**MethyLight.** The MethyLight assay is a high-throughput quantitative methylation assay that utilizes fluorescence-based real-time PCR (TaqMan ®) technology that requires no further manipulations after the PCR step (Eads et al., *Cancer Res.* 59:2302-2306, 1999).

Briefly, the MethyLight process begins with a mixed sample of genomic DNA that is converted, in a sodium bisulfite reaction, to a mixed pool of methylation-dependent sequence differences according to standard procedures (the bisulfite process converts unmethylated cytosine residues to uracil). Fluorescence-based PCR is then performed either in an "unbiased" (with primers that do not overlap known CpG methylation sites) PCR reaction, or in a "biased" (with PCR primers that overlap known CpG dinucleotides) reaction. Sequence discrimination can occur either at the level of the amplification process or at the level of the fluorescence detection process, or both.

The MethyLight assay may be used as a quantitative test for methylation patterns in the genomic DNA sample, wherein sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing of the biased PCR pool with either control oligonucleotides that do not "cover" known methylation sites (a fluorescence-based version of the "MSP" technique), or with oligonucleotides covering potential methylation sites.

The MethyLight process can be used with a "TaqMan®" probe in the amplification process. For example, double-stranded genomic DNA is treated with sodium bisulfite and subjected to one of two sets of PCR reactions using TaqMan® probes; e.g., with either biased primers and TaqMan® probe, or unbiased primers and TaqMan® probe. The TaqMan® probe is dual-labeled with fluorescent "reporter" and "quencher" molecules, and is designed to be specific for a relatively high GC content region so that it melts out at about 10 °C higher temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the

Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5' to 3' endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

Typical reagents (e.g., as might be found in a typical MethylLight-based kit) for MethylLight analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); TaqMan® probes; optimized PCR buffers and deoxynucleotides; and Taq polymerase.

**Ms-SNuPE.** The Ms-SNuPE technique is a quantitative method for assessing methylation differences at specific CpG sites based on bisulfite treatment of DNA, followed by single-nucleotide primer extension (Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997). Briefly, genomic DNA is reacted with sodium bisulfite to convert unmethylated cytosine to uracil while leaving 5-methylcytosine unchanged. Amplification of the desired target sequence is then performed using PCR primers specific for bisulfite-converted DNA, and the resulting product is isolated and used as a template for methylation analysis at the CpG site(s) of interest. Small amounts of DNA can be analyzed (e.g., microdissected pathology sections), and it avoids utilization of restriction enzymes for determining the methylation status at CpG sites. Typical reagents (e.g., as might be found in a typical Ms-SNuPE-based kit) for Ms-SNuPE analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE primers for specific gene; reaction buffer (for the Ms-SNuPE reaction); and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery regents or kit (e.g., precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

**MSP.** MSP (methylation-specific PCR) allows for assessing the methylation status of virtually any group of CpG sites within a CpG island, independent of the use of methylation-sensitive restriction enzymes (Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; US Patent No. 5,786,146). Briefly, DNA is modified by sodium bisulfite converting all unmethylated, but not methylated cytosines to uracil, and subsequently amplified with primers specific for methylated versus unmethylated DNA. MSP requires only small quantities of DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed on DNA extracted from paraffin-embedded samples. Typical reagents (e.g., as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation-altered DNA sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

**MCA.** The MCA technique is a method that can be used to screen for altered methylation patterns in genomic DNA, and to isolate specific sequences associated with these changes (Toyota et al., *Cancer Res.* 59:2307-12, 1999). Briefly, restriction enzymes with different sensitivities to cytosine methylation in their recognition sites are used to digest 5 genomic DNAs from primary tumors, cell lines, and normal tissues prior to arbitrarily primed PCR amplification. Fragments that show differential methylation are cloned and sequenced after resolving the PCR products on high-resolution polyacrylamide gels. The cloned fragments are then used as probes for Southern analysis to confirm differential methylation of these regions. Typical reagents (*e.g.*, as might be found in a typical MCA -based kit) for 10 MCA analysis may include, but are not limited to: PCR primers for arbitrary priming Genomic DNA; PCR buffers and nucleotides, restriction enzymes and appropriate buffers; gene-hybridization oligos or probes; control hybridization oligos or probes.

**Kits for Detection of Methylated CpG-containing Nucleic Acid.** The reagents required to perform one or more art-recognized methylation assays (including those identified 15 above) are combined with primers or probes comprising the sequences of SEQ ID NOS:1-103, or portions thereof, to determine the methylation state of CpG-containing nucleic acids. For example, the MethyLight, Ms-SNuPE, MCA, COBRA, and MSP methylation assays could be used alone or in combination, along with primers or probes comprising the 20 sequences of SEQ ID NOS:1-103, or portions thereof, to determine the methylation state of a CpG dinucleotide within a genomic sequence corresponding to SEQ ID NOS:1-103, or to CpG island sequences associated with sequences of SEQ ID NOS:1-103, where the CpG 25 island sequence associated with the sequence of the particular SEQ ID NO is that contiguous sequence of genomic DNA that encompasses at least one nucleotide of the particular SEQ ID NO sequence, and satisfies the criteria of having both a frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6, and a GC Content >0.5.